

SYSTEM AND METHOD FOR DYNAMICALLY ALLOCATING PROCESSING ON A NETWORK AMONGST MULTIPLE NETWORK SERVERS

Abstract of the Invention

The present invention provides a method, apparatus, and computer implemented instructions for
5 processing Web and other Internet or Intranet based services. The system for processing Web
requests includes a Web server with a connection to the Internet or Intranet with a predefined
network bandwidth, a set of primary Web and application server cluster nodes to process the
requests, and a dispatcher to allocate requests to nodes; in addition, one or more offload server
nodes are connected to the network. Client Web requests arrive at the dispatcher of the Web
10 server, which determines whether the incoming request can be handled at the primary Web server
cluster, whether all or part of the user Web request should be offloaded to one of the offload
server nodes, or whether the request should be throttled. If the dispatcher determines that the
request should be handled by the primary Web server cluster, it is appropriately routed to one of
the nodes in the primary Web server cluster; else if the dispatcher determines that the request
15 should be offloaded, one of the offload server nodes or service providers is selected, and the
request is either routed to a primary server node with the appropriate indication to offload all or
part of the request, or the request is routed to the selected offload service provider; otherwise, the
request is throttled by either routing it to a node which returns information that the service is
overloaded, or if the Web servers are too busy to provide even an overload indication, then the
20 request is dropped.